Analyzing Ordered Categorical Data derived from Elliptically Symmetric Distributions

Martin Kukuk *
University of Tübingen †

(First Version: December 10, 1998)

Abstract

The polychoric correlation is an ML estimator for the correlation parameter between two latent variables. Each latent variable is only observed as an ordered categorical indicator. This estimator is based on an assumption on the joint distribution for the latent variables which in this case is the bivariate standard normal distribution. We perform a simulation study applying the polychoric correlation based on normality if the true distribution is in fact an elliptically symmetric distribution. The results show that the polychoric correlation is robust in the sense that the true correlation between the latent variables is estimated only with small bias if the true distribution is not too leptokurtic and also not too platykurtic. These results imply that in practical applications the polychoric correlation can be applied obtaining meaningful results even if tests suggest that the assumed normal distribution is not appropriate. Basically the same results are obtained if one latent variable is observed directly and the ML-estimator based on normality (polyserial correlation) is applied.

KEY WORDS: Microeconometrics, Ordered Data, Latent Variables, Polychoric Correlation, Polyserial Correlation, Brillinger's Estimator, Elliptical Symmetric Distributions.

*Helpful comments on an earlier version from Robert Jung, Roman Liesenfeld, Gerd Ronning, and Peter von Tessin are gratefully acknowledged. The usual disclaimer applies.

†Address for correspondence: Wirtschaftswissenschaftliche Fakultät, Universität Tübingen, Mohlstr. 36, 72074 Tübingen, Germany. E-mail: Martin.Kukuk@uni-tuebingen.de.

1 Introduction

Business surveys or household surveys are often designed to obtain information on continuous variables which are hard to quantify exactly. Therefore, the questionnaire supplies ordered categorical answers like income categories, tendencies of change, degree of likes or dislikes, or degree of satisfaction. The scale of those variables is problematic if they ought to be used as explanatory variables in a regression analysis. One suggestion to deal with these ordinal data is to assume an underlying latent variable for each categorical indicator. This idea is widespread in various branches of applied statistical analysis like microeconometrics, biometrics, and psychometrics where ordinal data often occur.

The estimation of the linear dependency between two latent continuous variables, for which only ordered categorical observations are available, has a long tradition. Karl Pearson (1901) introduced this idea for a two-by-two contingency table. For each dichotomous indicator he assumed an underlying unobservable continuous variable. The latent continuous variables are assumed to jointly follow a standard bivariate normal distribution. The four frequencies of the two-by-two contingency table are used to estimate the correlation parameter of the latent continuous model. Pearson called this estimator *tetrachoric* correlation.

The bivariate normality assumption has two attractive properties: Firstly, the correlation is equal to the only parameter entering the standard bivariate normal distribution. Secondly, the regression of one variable on the other is linear and the regression parameter is equal to the correlation coefficient. This implies, especially in an econometric setting, that we can formulate a latent linear regression model, where the dependent variable and also the explanatory variable are measured categorically, and that we can obtain the regression parameter by estimating the parameter of the joint distribution.

Further development of this idea is closely related to advances in computing power. Tallis (1962) described the ML-estimation of the correlation parameter between the latent variables using two-by-two and also three-by-three contingency tables. The method involves iterative procedures to obtain the ML-estimate. Olsson (1979) extended the ML-estimation for general $r \times s$ contingency tables. Hamdan (1970) showed the equivalence of Pearson's tetrachoric correlation and Tallis' ML-estimator in the two-by-two case. The ML-estimator is usually called *polychoric correlation*.

Analogously, in a multivariate setting, where for each latent variable an ordinal indicator is observable, a standard multivariate normal distribution is assumed to estimate the correlation

matrix using the multiway contingency table (Poon and Lee, 1987, Kukuk, 1991). This latent model is attractive since the marginal distribution of a subset of latent variables given the others is again normal with conditional means being linear functions of the given variables. The regression parameters are again functions of the correlation matrix coefficients of the joint model. However, the calculation of probabilities for multivariate normal distributions is still a large computational burden and even intractable for higher dimensions. Kukuk (1991) performed some Monte-Carlo studies and showed that the efficiency loss of estimating the correlation matrix in a pair-wise fashion is negligible.

The distributional assumption for the latent variables can generally be tested using Pearson's X^2 test (Kukuk, 1991). In practical applications this assumption is often rejected (Kukuk, 1994). Therefore, Lee and Lam (1988) suggested ML-estimates using members of the elliptical distribution class containing the normal distribution as well as multivariate t-distributions. The problem is that the distribution has to be determined in advance in order to estimate the model. However, their simulation results indicate that applying the normality assumption in the estimation procedure although the latent variables follow a bivariate t-distribution leads to surprisingly small biases. Our focus is to investigate more members of the elliptical distribution class. Two subclasses are considered: firstly, the Kotz-type, where the multivariate normal distribution is a special case, and secondly the Pearson-type VII distributions, where the multivariate Cauchy distribution and t-distributions are special cases (Fang et al., 1990).

The multivariate elliptically symmetric distributions play an important role in another branch of microeconometrics. That research is concerned with generalizing results that binary dependent variable models, incorrectly estimated with OLS, still obtain (up to a scalar) consistent parameter estimates (e.g. Ruud, 1986, Stoker, 1986). The regressor variables are assumed to follow an elliptical distribution and, hence, the latent dependent variable and all the regressors follow a joint distribution which is in contrast to fixed regressors usually assumed in econometric modeling. In this sense it is analogous to our discussion that the "observed" latent variables are interpreted as a random sample of a joint distribution (Ruud, 1983). The regression function is deducted from this joint distribution. As a result, regression parameters are obtained via estimated parameters of the joint distribution (Ronning and Kukuk, 1996).

The paper is organized as follows. In the next section the spherically symmetric distribution class and the elliptically symmetric distribution class are introduced. The characterization of the distributions will be the basis for simulating these distributions. In section 3 estimation of

the polychoric correlation is discussed as well as the closely related *polyserial correlation* which is an ML-estimator in the situation where one latent continuous variable is observed directly. Those estimators are used in the simulation study in section 4 to study how sensitive they are if the underlying distribution deviates from normality. Section 5 summarizes and concludes.

2 Multivariate Spherical and Elliptical Distributions

In this section a distribution class is considered which generalizes the multivariate normal distribution in the sense that the above outlined attractive features are preserved. The probability density function of a multivariate normal distribution is given by

$$\phi_n(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{\frac{1}{2}n} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left(\Leftrightarrow \frac{1}{2} (\boldsymbol{x} \Leftrightarrow \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} \Leftrightarrow \boldsymbol{\mu}) \right)$$

where $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)'$, $\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{x})$, and $\boldsymbol{\Sigma} = \mathrm{E}(\boldsymbol{x} \Leftrightarrow \boldsymbol{\mu})(\boldsymbol{x} \Leftrightarrow \boldsymbol{\mu})'$. In case the covariance matrix is equal to the identity matrix $\boldsymbol{I}$ and $\boldsymbol{\mu} = \boldsymbol{0}$ the density function can be written as

$$\phi_n(\boldsymbol{x}) \propto \exp(\Leftrightarrow \boldsymbol{x}'\boldsymbol{x}/2) \quad ,$$

where the variables $x_1, x_2, \ldots, x_n$ only enter the density function through the inner product $\boldsymbol{x}'\boldsymbol{x}$. It implies that all elements $\boldsymbol{x} \in \mathbb{R}^n$ having the same Euclidean distance from the origin have the same value of the density function ϕ_n. In other words, contours of surfaces of equal density are *spheres* around the origin with radius $r = (\boldsymbol{x}'\boldsymbol{x})^{1/2}$. This concept is used to define a distribution class of *spherically symmetric distributions* (or simply spherical distributions) having probability density functions

$$\psi(\boldsymbol{x}) \;=\; h(\boldsymbol{x}'\boldsymbol{x}) \quad . \tag{1}$$

This distribution class is only a subset of a broader class of spherically symmetric distributions since it requires that the distributions possess densities. The broader class can be defined using a stochastic representation (Fang et al., 1990)

$$\boldsymbol{x} \stackrel{d}{=} r \cdot \boldsymbol{u} \quad , \tag{2}$$

where $\stackrel{d}{=}$ signifies that both sides have the same distribution, $\boldsymbol{u}$ being a random vector uniformly distributed on the unit sphere, and r is a positive random variable independent of $\boldsymbol{u}$. This stochastic representation will be used later on to obtain pseudo random variables following a spherical distribution.

The relationship between the density $h(\cdot)$ and the density of random variable r can be established using the condition

$$\int_{\mathbb{R}^n} h(\boldsymbol{x}'\boldsymbol{x})\, d\boldsymbol{x} = 1 \tag{3}$$

and transforming the variables $x_1, x_2, \ldots, x_n$ to polar coordinates $r, \theta_1, \theta_2, \ldots, \theta_{n-1}$ (Anderson, 1984 p.279):

$$\begin{aligned}
y_1 &= r\sin\theta_1 \quad , \\
y_2 &= r\cos\theta_1\sin\theta_2 \quad , \\
&\vdots \\
y_{n-1} &= r\cos\theta_1\cos\theta_2\cdots\cos\theta_{n-2}\sin\theta_{n-1} \quad , \\
y_n &= r\cos\theta_1\cos\theta_2\cdots\cos\theta_{n-2}\cos\theta_{n-1} \quad .
\end{aligned}$$

The Jacobian of this transformation is $r^{n-1}\cos^{n-2}\theta_1\cos^{n-2}\theta_2\cdots\cos\theta_{n-2}$. Therefore, Equation (3) can be written as

$$\int_{\mathbb{R}^n} h(\boldsymbol{x}'\boldsymbol{x})\, d\boldsymbol{x} = \int_0^\infty \int_{\boldsymbol{\Theta}} h(r^2)\cdot r^{n-1}\cos^{n-2}\theta_1\cos^{n-2}\theta_2\cdots\cos\theta_{n-2}\, d\boldsymbol{\theta} dr = 1$$

where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{n-1})'$. The multiple integral can be solved recursively using

$$\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^{m-1}\varphi\, d\varphi = \frac{\Gamma(\frac{1}{2}m)\Gamma(\frac{1}{2})}{\Gamma\left(\frac{1}{2}(m+1)\right)}$$

resulting in

$$\int_{\mathbb{R}^n} h(\boldsymbol{x}'\boldsymbol{x})\, d\boldsymbol{x} = \frac{2\pi^{n/2}}{\Gamma(\frac{1}{2}n)}\int_0^\infty h(r^2)\cdot r^{n-1}\, dr = 1 \quad .$$

Substituting $y = r^2$ implying $dr = (2r)^{-1}dy$, the following relation is obtained:

$$\int_{\mathbb{R}^n} h(\boldsymbol{x}'\boldsymbol{x})\, d\boldsymbol{x} = \frac{\pi^{n/2}}{\Gamma\left(\frac{1}{2}n\right)}\int_0^\infty h(y)\cdot y^{n/2-1}\, dy = 1 \quad . \tag{4}$$

The function $h(\cdot)$ is called the *density generator* of a spherical distribution if random variable r in Equation (2) has a density $f(\cdot)$ with (Fang et al, 1990 p. 35)

$$f(r) = \frac{2\pi^{n/2}}{\Gamma\left(\frac{1}{2}n\right)} r^{n-1} h(r^2) \quad . \tag{5}$$

As an example, if $h(\boldsymbol{x}'\boldsymbol{x}) = (2\pi)^{-n/2}\exp(\Leftrightarrow\boldsymbol{x}'\boldsymbol{x}/2)$ which is the joint density of n independently and standard normally distributed variables, we derive from (4) the density of $q = \boldsymbol{x}'\boldsymbol{x}$ as

$$\begin{aligned} g(q) &= \frac{\pi^{n/2}}{\Gamma\left(\frac{n}{2}\right)}\frac{1}{(2\pi)^{n/2}}\exp\left(\Leftrightarrow\frac{q}{2}\right)q^{n/2-1} \\ &= \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)}\exp\left(\Leftrightarrow\frac{q}{2}\right)q^{n/2-1} \end{aligned}$$

which is the density of a χ^2_n-distribution establishing the well-known fact that the sum of n squared independently and standard normally distributed variables $q = \sum_{i=1}^{n} x_i^2 = \boldsymbol{x}'\boldsymbol{x}$ follows a chi-square distribution with n degrees of freedom. Equation (5) yields the density of r which is the chi-distribution with n degrees of freedom (Johnson et al., 1994 p. 417)

$$\begin{aligned} f(r) &= \frac{2\pi^{n/2}}{\Gamma\left(\frac{1}{2}n\right)}r^{n-1}\frac{1}{(2\pi)^{n/2}}\exp\left(\Leftrightarrow\frac{r^2}{2}\right) \\ &= \frac{2^{n/2}}{\Gamma\left(\frac{1}{2}n\right)2^{n/2-1}}r^{n-1}\exp\left(\Leftrightarrow\frac{r^2}{2}\right) \quad . \end{aligned}$$

In order to obtain a spherical distribution with density generator $h(\cdot)$ and hence density $f(r)$ according to Equation (2) a random vector $\boldsymbol{u}$ is required being uniformly distributed on the unit sphere. Such a vector can be constructed using a random vector $\boldsymbol{y}$ following a multivariate normal distribution with density $\phi_n(\boldsymbol{y}, \boldsymbol{0}, \boldsymbol{I})$. Defining

$$u_i = \frac{y_i}{\|\boldsymbol{y}\|} \qquad\qquad i = 1, \ldots, n \tag{6}$$

where $\|\boldsymbol{y}\| = (\boldsymbol{y}'\boldsymbol{y})^{1/2}$ then $\boldsymbol{u}$ is uniformly distributed on the unit sphere in $\mathbb{R}^n$. This procedure is convenient since standard normal (pseudo) random number generators are implemented in many software packages. Thus, simulating $\boldsymbol{u}$ and the univariate random variable r independently of $\boldsymbol{u}$ and multiplying them according to (2) yields (pseudo) random numbers of spherical distributions with density $h(\boldsymbol{x}'\boldsymbol{x})$. r will be simulated in our study using the Acceptance-Rejection Method. Analogously to the multivariate normal distribution where the general distribution can be derived by the spherical normal distribution using the linear transformation

$$\boldsymbol{z} = \boldsymbol{\mu} + \boldsymbol{A}'\boldsymbol{x}$$

with $\boldsymbol{\Sigma} = \boldsymbol{A}'\boldsymbol{A}$, $\mathrm{rank}(\boldsymbol{\Sigma}) = n$, and $\boldsymbol{\mu} \in \mathbb{R}^n$. The contours of surfaces of equal density are now *ellipsoids* around the center $\boldsymbol{\mu}$. The covariance matrix for the general normal distribution is equal to $\boldsymbol{\Sigma}$. We use the same linear transformation for any $\boldsymbol{x}$ following a spherical distribution and obtain a class of *elliptically symmetric* distributions. For this class it can be shown that the covariance matrix of $\boldsymbol{z}$, if those moments exist, is proportional to $\boldsymbol{\Sigma}$. For the partitioned vector

$\boldsymbol{z} = (\boldsymbol{z}^{(1)}, \boldsymbol{z}^{(2)})'$ with appropriate

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \end{pmatrix} \qquad \text{and} \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}' & \boldsymbol{\Sigma}_{22} \end{pmatrix}$$

the distribution of $\boldsymbol{z}^{(1)} \mid \boldsymbol{z}_0^{(2)}$ is again elliptical with $\boldsymbol{\mu}_{1.2} = \boldsymbol{\mu}^{(1)} + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{z}_0^{(2)} \Leftrightarrow \boldsymbol{\mu}^{(2)})$. In other words the regression of a subset of variables on the others is a linear function (Fang et al., 1990 p. 45).

3 Estimators for Categorical Data

Polychoric Correlation

As mentioned above, the covariance matrix of a multivariate distribution can be estimated in a pair-wise fashion, especially when the multivariate treatment is computational burdensome or intractable. Therefore, we want to analyze the bivariate situation where for the two latent variables z_1^* and z_2^* only categorical indicators are available:

$$z_i = j \qquad \Leftrightarrow \qquad \alpha_{i;j-1} \leq z_i^* < \alpha_{i;j} \qquad\qquad j = 1, 2, \ldots, k_i \quad i = 1, 2 \quad . \tag{7}$$

The first and last threshold α_0 and α_{k_i} of each variable is equal to $\Leftrightarrow\infty$ and ∞, respectively. The observations for the categorical indicators z_1 and z_2 from a random sample can be summarized in a $k_1 \times k_2$ contingency table containing the relative or absolute frequencies. Assuming a multivariate normal distribution for z_1^* and z_2^* we see from measurement relation (7) that location and scale parameters of the latent variables are not identified unless we impose restrictions on the thresholds. Thus, denoting the standard bivariate normal distribution function by $\Phi_2(\cdot, \cdot, \rho)$, the probability of observation $z_1 = i \wedge z_2 = j$ can be written as

$$\pi_{ij} = \Phi_2(\alpha_{1;i}, \alpha_{2;j}, \rho) \Leftrightarrow \Phi_2(\alpha_{1;i-1}, \alpha_{2;j}, \rho) \Leftrightarrow \Phi_2(\alpha_{1;i}, \alpha_{2;j-1}, \rho) + \Phi_2(\alpha_{1;i-1}, \alpha_{2;j-1}, \rho) \quad .$$

Maximizing the Log–Likelihood function

$$L(\rho, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2) = \sum_{i=1}^{k_1} \sum_{j=1}^{k_2} n_{ij} \log \pi_{ij} \quad ,$$

where n_{ij} is the absolute frequency in the i, j cell of the contingency table, the polychoric correlation $\hat{\rho}$ is obtained together with estimates of the thresholds (Olsson, 1979; Kukuk, 1991). From simulation studies it is known that sample sizes of $N = 100$ for 3×3 contingency tables yield sufficiently small biases for ρ unless the thresholds are chosen so that observations in some

categories are highly unlikely. Even in situations where the thresholds are chosen so that the distributions of the categorical indicators are highly skewed the bias is still in the same negligible order (Olsson, 1979; Poon and Lee, 1987, Kukuk, 1991). Thus, if the joint distribution of z_1^* and z_2^* is correctly assumed, the estimation procedure works well. Pearson's X^2 test can be used in this setting to evaluate whether the assumed distribution of the latent variables is appropriate. The test statistic is

$$X^2 = \sum_{i=1}^{k_1} \sum_{j=1}^{k_2} \frac{(n_{ij} \Leftrightarrow N\pi_{ij})^2}{N\pi_{ij}} \quad ,$$

which is χ^2_ν distributed with $\nu = (k_1 \cdot k_2 \Leftrightarrow k_1 \Leftrightarrow k_2)$ degrees of freedom under the null hypothesis of the correct distribution assumption. Analyzing real data, the normality assumption is often strongly rejected. Lee and Lam (1988) tackled this problem by choosing other elliptical distributions in the ML procedure. However, the first difficulty is to choose the most appropriate member and the second one is that the distribution function is in most cases not available in a closed analytical form. Therefore, we want to perform a simulation study using different elliptical distributions but still using the ML procedure based on the normality assumption to estimate the correlation ρ. A first indication of the robustness of the latter procedure are the results reported in Lee and Lam (1988). Besides the estimates obtained with ML based on the correct t-distribution they showed the estimates from ML based on normality which are quite close.

It should be mentioned that since the continuous variables are not observed directly monotone transformations of z_i^* and correspondingly α_i result in unchanged observations of z_i. Therefore, for all members of Mardia's distribution class (Mardia, 1970) we observe the same contingency table and as a consequence obtain the same correlation estimate (Kukuk, 1994). However, for our purposes this distribution class is not too interesting since the regression function of one latent variable on the other is mostly not linear.

Polyserial Correlation and Brillinger's Estimator

Olsson et al. (1982) developed an ML procedure for the situation where the latent variables z_1^* and z_2^* again follow a bivariate normal distribution but only one variable z_1 is observed categorically and z_2^* is directly observable. They named the estimator *polyserial correlation.* In this case Brillinger's (1982) one-step estimator can be applied. His main result states that under

mild conditions for a measurable function $s(\cdot)$

$$\text{cov}(z_1^*, z_2^*) = \frac{\text{cov}(s(z_1^*), z_2^*)\text{var}(z_1^*)}{\text{cov}(s(z_1^*), z_1^*)}$$

which in our case with $s(\cdot)$ given in Equation (7) leads to

$$\tilde{\rho}_{z_1^* z_2^*} = \frac{\rho^{BP}_{z_1 z_2^*}}{\breve{\rho}_{z_1^* z_1}} \quad , \tag{8}$$

where ρ^{BP} denotes the Bravais-Pearson correlation coefficient using observations for z_1 and z_2^* and

$$\begin{aligned}
\breve{\rho}_{z_1^* z_1} &= \frac{\text{E}(z_1 \cdot z_1^*)}{(\text{Var}(z_1))^{1/2}} = \frac{\text{E}\left[\text{E}(z_1^*|z_1)\right]}{\left[\sum_{j=1}^{k_1} j^2 \cdot (\Phi(\alpha_j) \Leftrightarrow \Phi(\alpha_{j-1})) \Leftrightarrow \left(\sum_{j=1}^{k_1} j \cdot (\Phi(\alpha_j) \Leftrightarrow \Phi(\alpha_{j-1}))\right)^2\right]^{-1/2}} \\
&= \sum_{j=1}^{k_1} \phi_1(\alpha_j) \left[k_1^2 \Leftrightarrow \sum_{j=1}^{k_1-1} (2j+1) \cdot \Phi_1(\alpha_j) \Leftrightarrow \left(k_1 \Leftrightarrow \sum_{j=1}^{k_1-1} \Phi_1(\alpha_j)\right)^2\right]^{-1/2}
\end{aligned}$$

with $\phi_1(\cdot)$ and $\Phi_1(\cdot)$ denoting the density and distribution function of the univariate standard normal distribution, respectively. Brillinger's estimator is very easy to calculate requiring only first-step estimates of the thresholds which are given by

$$\hat{\alpha}_{1;j} = \Phi_1^{-1}\left(\sum_{l=1}^{j} \sum_{m=1}^{k_2} \frac{n_{lm}}{N}\right) \qquad j = 1, \ldots, k_1 \Leftrightarrow 1 \quad .$$

In the following simulation study the polychoric correlation, the polyserial correlation as well as Brillinger's estimator, all relying on the normality assumption, will be analyzed in non-normal situations.

4 Monte-Carlo Study

In this section we simulate two standardized continuous variables jointly following an elliptical distribution allowing the marginal distributions of the latent variables to be leptokurtic or platykurtic. As mentioned earlier, for all members of this distribution class the parameter of interest ρ is the correlation between the latent variables and it equals the regression parameter of one latent variable on the other. We can restrict the analysis to two variables since the correlation matrix $\boldsymbol{R}$ in the multivariate case can always be estimated in a pair-wise fashion (Kukuk,1991).

The general setup is that we first use a large sample of 100.000 observations to study the general behaviour of the estimators. This will be done with 3 categories per variable since in the case of

a 2×2 contingency table the number of parameters equals the number of free frequencies and therefore the X^2 test is not applicable. Firstly, we use sets of symmetric thresholds resulting in symmetric discrete distributions of the categorical variables. Secondly, the thresholds are chosen to obtain highly skewed discrete distributions. For each continuous distribution the sets of thresholds are determined to obtain the same categorical distributions in all settings. In a next step these large samples are each divided in 100 samples of 1.000 observations to analyze the small sample properties. This is the sample size we often encounter in practical applications. The results shown in the following are all for $\rho = \Leftrightarrow .8$. For other values the results are basically the same and are therefore not reported.

Symmetric Kotz-Type Distributions

The density functions of standardized bivariate elliptical Kotz-type distributions are given by (Fang et al., 1990, p. 76)

$$h(\boldsymbol{x}) = c \cdot |\boldsymbol{R}|^{-1/2} \left(\boldsymbol{x}'\boldsymbol{R}^{-1}\boldsymbol{x}\right)^{N-1} \exp\left(\Leftrightarrow r \cdot \left(\boldsymbol{x}'\boldsymbol{R}^{-1}\boldsymbol{x}\right)^{s}\right) \qquad N, r, s > 0 \quad ,$$

where c is the normalizing constant and $\boldsymbol{R}$ the correlation matrix. For $N = 1$, $s = 1$, and $r = 1/2$ we obtain the bivariate normal distribution. For $N < 1$ the density function tends to infinity at the origin, whereas for $N > 1$ the density function has a local minimum at the origin and looks like a volcano crater. For $N = 1$, the density function appears more and more cylindrical as s grows larger. In the Monte Carlo simulations we vary N from 0.1 to 3 and s from 0.25 to 8, whereas r is kept fixed at $1/2$. The polychoric correlations for large samples of 100.000 observations are recorded in table 1 for various sets of thresholds. First of all it can be seen that in the situation $N = 1$ and $s = .9$, which is close to the normal distribution, the estimator works well for all threshold combinations.

The polychoric correlation works also well for all continuous distributions considered if both categorical variables have almost equal probabilities (columns .3/.4 and .35/.3). If the outer categories have smaller frequencies (column .1/.8), which is, for instance, more realistic for categorical data in business surveys, the estimator obtains adequate results if the distributions are not too leptokurtic (empirical kurtosis $\hat{\sigma}^4$ of the latent variables less than 20) and also not too platykurtic ($\hat{\sigma}^4 > 2.5$). Outside this range, the true correlation is underestimated in absolute terms. However, for this threshold setting there is a tendency of overestimating ρ when the empirical kurtosis is around 4-6.

Table 1: Polychoric Correlation for Kotz-Type Distributions ($\rho = \Leftrightarrow.8$). Large Samples.

			Marginal frequencies of $z_1 = 1/z_1 = 2$ $z_2 = 1/z_2 = 2$								
N	s	$\hat{\sigma}^4$	.1/.8	.15/.7	.2/.6	.25/.5	0.3/.4	.35/.3	.45/.35	.45/.35	.45/.35
			.1/.8	.15/.7	.2/.6	.25/.5	0.3/.4	.35/.3	.5/.35	.3/.4	.15/.35
.1	.3	47	-.752	-.749	-.752	-.757	-.768	-.783	-.705	-.744	-.775
.1	.5	14	-.806	-.785	-.775	-.775	-.782	-.790	-.733	-.763	-.779
.1	1.5	4.1	-.829	-.821	-.810	-.802	-.797	-.797	-.774	-.784	-.789
.1	8	3.0	-.810	-.819	-.826	-.811	-.801	-.798	-.785	-.789	-.794
.5	.25	26	-.777	-.762	-.760	-.765	-.774	-.784	-.711	-.742	-.773
.5	.9	4.7	-.835	-.819	-.803	-.794	-.791	-.795	-.765	-.776	-.787
.5	4	2.9	-.803	-.818	-.825	-.814	-.799	-.796	-.787	-.788	-.794
.5	8	2.75	-.800	-.816	-.824	-.819	-.805	-.799	-.790	-.792	-.794
.9	.25	15.5	-.809	-.792	-.782	-.781	-.784	-.789	-.740	-.763	-.785
.9	.7	4.0	-.825	-.815	-.804	-.797	-.795	-.796	-.785	-.791	-.797
.9	1.5	2.65	-.787	-.791	-.796	-.803	-.804	-.803	-.806	-.802	-.801
.9	2	2.45	-.773	-.782	-.794	-.801	-.805	-.804	-.813	-.807	-.805
1	.25	13	-.804	-.791	-.784	-.784	-.787	-.793	-.747	-.770	-.789
1	.5	5	-.828	-.816	-.802	-.793	-.791	-.795	-.777	-.785	-.793
1	.9	3.2	-.806	-.801	-.804	-.802	-.800	-.800	-.796	-.798	-.800
1	2	2.3	-.761	-.777	-.786	-.799	-.808	-.808	-.820	-.811	-.807
1.1	.25	12	-.812	-.795	-.785	-.784	-.787	-.791	-.751	-.772	-.787
1.1	.9	3.1	-.801	-.798	-.800	-.800	-.802	-.803	-.802	-.805	-.803
1.1	1.5	2.5	-.767	-.777	-.789	-.799	-.804	-.808	-.817	-.812	-.805
1.1	2	2.3	-.757	-.768	-.780	-.794	-.806	-.807	-.822	-.812	-.805
2	.25	6	-.818	-.803	-.793	-.789	-.788	-.791	-.776	-.785	-.794
2	.5	3.1	-.801	-.796	-.791	-.789	-.793	-.798	-.804	-.804	-.803
2	.7	2.6	-.777	-.777	-.780	-.785	-.797	-.802	-.816	-.813	-.805
2	.9	2.3	-.762	-.763	-.769	-.782	-.796	-.805	-.822	-.815	-.805
3	.25	4.2	-.813	-.801	-.790	-.788	-.790	-.798	-.792	-.797	-.802
3	.3	3.55	-.808	-.798	-.791	-.789	-.793	-.798	-.799	-.802	-.801
3	.5	2.55	-.775	-.769	-.773	-.781	-.791	-.801	-.817	-.813	-.805
3	.7	2.27	-.745	-.746	-.758	-.772	-.791	-.804	-.827	-.819	-.802

Note: $\hat{\sigma}^4$ denotes the empirical kurtosis of the standardized latent variables.

Table 2: X^2 from Polychoric Correlation for Kotz-Type Distributions ($\rho = \Leftrightarrow 8$)

			Marginal frequencies of $z_1 = 1/z_1 = 2$ $z_2 = 1/z_2 = 2$				
N	s	$\hat{\sigma}^4$	.1/.8 .1/.8	.2/.6 .2/.6	.3/.4 .3/.4	.35/.3 .35/.3	.45/.35 .15/.35
.1	.3	47	264030.03	37568.19	17850.33	13042.92	8206.03
.1	.5	14	334161.22	18216.45	9120.95	7810.40	4177.81
.1	1.5	4.1	163.93	2234.19	3733.68	4634.99	903.31
.1	8	3.0	3.05	3.96	2395.22	3847.67	257.87
1	.25	13	250491.63	14435.26	7923.01	6201.14	3479.11
1	.5	5	15127.57	2492.87	1604.08	1251.26	873.04
1	.9	3.2	10.05	26.27	40.68	27.77	13.76
1	2	2.3	1.98	104.75	551.79	458.76	458.61
2	.25	6	40273.65	2631.89	970.79	406.29	800.37
2	.5	3.1	6.20	0.52	148.97	406.06	14.60
2	.7	2.6	3.04	77.84	730.42	1418.10	290.00
2	.9	2.3	1.53	132.56	1195.12	2264.73	740.77

Note: The ($\alpha = 5\%$) critical value for the analyzed situations is 7.8. Larger X^2 statistics imply a rejection of the assumed bivariate normal distribution.

Skewed distributions of the categorical indicators do not worsen the good performance of the polychoric correlation which can be seen from the last column of table 1. In this case the first indicator has 45% of observations in the first category whereas the second indicator has 50% in the third category. For all distributions the results are acceptable. This result holds for positive correlations if the indicators are skewed in the same category. On the other side, in the situation where both indicators have high frequencies in the first category (column .45/.35 & .5/.35) the performance is not satisfying. In this extreme case of skewness acceptable estimates are obtained if the kurtosis of the latent variable is between 2.2 and about 6.

In table 2 Pearson's X^2 statistics, testing the bivariate normality assumption, are recorded for some of the settings shown in table 1. Those settings not shown in the table reveal qualitatively the same conclusions. The last three columns are those situations where the polychoric correlation works well. However, the test statistics are well above the 5% critical value of 7.8 for a 3×3 contingency table. On the other hand, we observe low values of the test statistic for platykurtic distributions and wide thresholds where the estimator performs poorly. As a consequence, the X^2 test statistic does not provide valuable information about the performance of the estimator. At least for the .1/.8/.1 frequency distribution of the categorical indicators it also implies that this criterion cannot be used to discriminate between distributions for the latent variables.

Table 3: Correlation Estimates for Kotz-Type Distributions ($\rho = \Leftrightarrow .8$). Small Samples.

Cat.	PC_L	BP		PC_S		PS		BRI	
		$\bar{\rho}$	s_ρ	$\bar{\rho}$	s_ρ	$\bar{\rho}$	s_ρ	$\bar{\rho}$	s_ρ
$N = .1 \;\; s = .3 \;\; \hat{\sigma}^4 = 47$									
.1/.8	-.752	-.800	.046	-.752	.040	-.697	.030	-.740	.044
.2/.6	-.751	-.800	.046	-.752	.028	-.712	.029	-.564	.036
.35/.3	-.783	-.800	.046	-.782	.021	-.751	.038	-.466	.031
$N = .1 \;\; s = 1.5 \;\; \hat{\sigma}^4 = 4.1$									
.1/.8	-.829	-.801	.014	-.830	.022	-.785	.020	-.838	.022
.2/.6	-.810	-.801	.014	-.811	.021	-.797	.016	-.800	.016
.35/.3	-.797	-.801	.014	-.797	.019	-.803	.016	-.760	.017
$N = 1 \;\; s = .25 \;\; \hat{\sigma}^4 = 13$									
.1/.8	-.804	-.799	.023	-.805	.037	-.752	.026	-.833	.032
.2/.6	-.784	-.799	.023	-.785	.025	-.775	.023	-.725	.024
.35/.3	-.793	-.799	.023	-.793	.018	-.800	.024	-.653	.025
$N = 1 \;\; s = .5 \;\; \hat{\sigma}^4 = 5$									
.1/.8	-.828	-.800	.015	-.829	.031	-.783	.022	-.835	.024
.2/.6	-.802	-.800	.015	-.802	.023	-.794	.019	-.785	.019
.35/.3	-.795	-.800	.015	-.795	.019	-.803	.018	-.749	.018
$N = 1 \;\; s = .9 \;\; \hat{\sigma}^4 = 3.2$									
.1/.8	-.806	-.800	.013	-.806	.027	-.798	.021	-.806	.022
.2/.6	-.804	-.800	.013	-.804	.021	-.800	.017	-.800	.018
.35/.3	-.800	-.800	.013	-.800	.019	-.801	.017	-.795	.017
$N = 2 \;\; s = .25 \;\; \hat{\sigma}^4 = 6$									
.1/.8	-.818	-.799	.015	-.820	.029	-.775	.021	-.827	.023
.2/.6	-.793	-.799	.015	-.794	.025	-.790	.021	-.772	.020
.35/.3	-.791	-.799	.015	-.791	.020	-.800	.019	-.737	.018
$N = 2 \;\; s = .9 \;\; \hat{\sigma}^4 = 2.3$									
.1/.8	-.761	-.800	.010	-.761	.028	-.809	.019	-.758	.018
.2/.6	-.769	-.800	.010	-.770	.019	-.798	.014	-.790	.015
.35/.3	-.805	-.800	.010	-.805	.015	-.796	.011	-.829	.014
$N = 2 \;\; s = 3 \;\; \hat{\sigma}^4 = 1.8$									
.1/.8	-.668	-.800	.009	-.668	.037	-.816	.016	-.710	.016
.2/.6	-.731	-.800	.009	-.731	.023	-.795	.013	-.781	.016
.35/.3	-.810	-.800	.009	-.810	.014	-.789	.010	-.851	.014

Note: PC_L is polychoric correlation from Large Sample.
Bravais-Pearson $BP_{z_1^* z_2^*}$, polychoric PC_S, polyserial PS,
and Brillinger BRI are calculated from Small Samples.

Splitting the large samples each into 100 small samples with 1.000 observations we can study whether an extra bias occurs in small samples and also obtain standard deviations. In table 3 some parameter combinations already discussed in table 1 are considered. Comparing the mean of the polychoric correlations of the 100 samples $\bar{\rho}$ (PC_S) with the large sample estimates (recorded again in column PC_L) indicates that no small sample bias occurs.

Using one categorical variable z_1 and the latent variable z_2^*, implying that the latter can be observed directly, we estimate the polyserial correlation (PS) and Brillinger's estimator (BRI). The polyserial correlation performs better if the categories are distributed evenly (third row of each panel) just like the polychoric correlation. Unlike the latter, PS shows a bias when the latent variables have kurtosis greater than 20. For the other threshold settings the polyserial correlation works well for kurtosis less than 6.

The polyserial correlation is robust if the true distribution is platykurtic even in those situations where the polychoric correlation deviates more from the true value. On the contrary, Brillinger's estimator depends heavily on the normality assumption. The bias occurring with other elliptical distributions is in some settings substantial. No threshold values can be identified for which this estimator performs best. Thus, there is a clear dominance of the polyserial correlation over Brillinger's estimator.

The estimators' standard deviations are reported as well. For most settings the ordering of the estimators according to their standard deviation is as expected. Estimating the correlation using the continuous variables (BP) usually results in a smaller variance than only having one continuous variable and one categorical indicator (PS) or even both variables being categorical (PC). However, for those elliptical distributions having large empirical kurtosis the ordering of estimators according to their standard deviation is $s_{BP} > s_{PS} > s_{PC}$. This occurs whenever the kurtosis is greater than approximately 10.

Symmetric Pearson-Type VII Distributions

As a second subclass of elliptical distributions, we analyze the Pearson-type VII distributions which are characterized by the density generator

$$h(u) = c \cdot \left(1 + \frac{u}{m}\right)^{-N} \qquad N > 1,\ m > 0 \quad .$$

The multivariate *Cauchy distribution* is a member of this subclass having the parameters $N = 1.5$ and $m = 1$. If the parameters obey $N = 1+m/2$ we obtain multivariate *t-distributions*. The shape

Table 4: Correlation Estimates for Pearson-Type VII Distributions ($m = 1 \quad \rho = \Leftrightarrow 8$). Small Samples.

Cat.	PC_L	BP		PC_S		PS		BRI	
		$\bar{\rho}$	s_ρ	$\bar{\rho}$	s_ρ	$\bar{\rho}$	s_ρ	$\bar{\rho}$	s_ρ
				$N = 1.5 \;\; \hat{\sigma}^4 = 118.8$					
.1/.8	-.721	-.796	.068	-.725	.047	-.611	.048	-.544	.058
.2/.6	-.749	-.796	.068	-.749	.030	-.636	.044	-.395	.044
.35/.3	-.783	-.796	.068	-.783	.020	-.706	.061	-.325	.035
				$N = 1.8 \;\; \hat{\sigma}^4 = 89.8$					
.1/.8	-.766	-.796	.062	-.767	.042	-.707	.029	-.675	.060
.2/.6	-.765	-.796	.062	-.765	.025	-.746	.024	-.545	.050
.35/.3	-.787	-.796	.062	-.787	.022	-.815	.025	-.482	.047
				$N = 2.2 \;\; \hat{\sigma}^4 = 32.7$					
.1/.8	-.797	-.800	.034	-.801	.038	-.750	.029	-.778	.036
.2/.6	-.780	-.800	.034	-.781	.021	-.786	.019	-.684	.037
.35/.3	-.798	-.800	.034	-.798	.020	-.828	.022	-.636	.037
				$N = 2.4 \;\; \hat{\sigma}^4 = 20.5$					
.1/.8	-.802	-.800	.031	-.805	.036	-.760	.028	-.802	.031
.2/.6	-.784	-.800	.031	-.784	.022	-.791	.021	-.715	.029
.35/.3	-.792	-.800	.031	-.792	.020	-.819	.023	-.670	.032
				$N = 2.7 \;\; \hat{\sigma}^4 = 13.1$					
.1/.8	-.808	-.802	.023	-.810	.032	-.770	.025	-.811	.027
.2/.6	-.790	-.802	.023	-.791	.023	-.795	.020	-.744	.022
.35/.3	-.796	-.802	.023	-.796	.018	-.816	.022	-.707	.025
				$N = 3 \;\; \hat{\sigma}^4 = 9.7$					
.1/.8	-.812	-.801	.019	-.814	.034	-.772	.022	-.813	.023
.2/.6	-.788	-.801	.019	-.789	.025	-.793	.022	-.756	.021
.35/.3	-.795	-.801	.019	-.795	.019	-.809	.024	-.724	.021
				$N = 5 \;\; \hat{\sigma}^4 = 4.4$					
.1/.8	-.808	-.801	.014	-.809	.026	-.788	.021	-.813	.020
.2/.6	-.798	-.801	.014	-.798	.022	-.797	.018	-.788	.017
.35/.3	-.799	-.801	.014	-.800	.017	-.804	.016	-.773	.016
				$N = 10 \;\; \hat{\sigma}^4 = 3.2$					
.1/.8	-.798	-.800	.012	-.798	.025	-.796	.017	-.807	.019
.2/.6	-.813	-.800	.012	-.813	.022	-.804	.016	-.807	.017
.35/.3	-.798	-.800	.012	-.798	.016	-.802	.014	-.791	.015

Note: See table 3 for details.

of the spherical density function is always similar to the shape of the spherical normal distribution. As an example, starting from the Cauchy distribution and increasing m or decreasing N, the local maximum at the origin lowers and the outer regions rise, but the typical hill shape is still present.

Table 4 shows in the first panel results for the Cauchy distribution and the following panels show the results for distributions varying the parameter N while keeping $m = 1$ constant. The conclusions are in accordance with the discussion of Kotz-type distributions. The results can be summarized as follows:

- The polychoric correlations performs best whenever the frequencies of the discrete indicators are distributed almost uniformly. Even for the Cauchy distribution the large sample bias of the polychoric correlation is about 0.017. Increasing the value of parameter N reduces the empirical kurtosis. As a result, the estimates for the other threshold combinations improve and become acceptable.

- For the threshold values yielding .1/.8/.1 distributions for the categorical indicators the polychoric correlation (absolutely) overestimates ρ when the kurtosis is of magnitude 4 to 10.

- Small samples of 1.000 observations yield the same estimates as large samples. The standard deviation of the polychoric correlation is smaller than the one of the Bravais-Pearson correlation coefficient if the kurtosis is greater than 10 (.35/.3/.35 frequencies) or greater than 20 (.15/.7/.15 frequencies).

- The polyserial correlation performs best with distributions having lower values of kurtosis. For kurtosis values less than 8 the estimator shows only small biases for all threshold settings. For distributions with higher kurtosis good results are obtained only if the discrete distributions of the categorical indicator show almost even frequencies. The polyserial correlation outperforms Brillinger's estimator which can be recommended only for small values of excess kurtosis.

The t-distributions require the parameters to be $N = 1 + m/2$. In this sense the Cauchy distribution can be interpreted as a t-distribution. We simulated other t-distributions, which are not reported in the tables, starting with $N = 2$ and $m = 2$. Already for this setting the bias of the polychoric correlation is less than 0.02 for all the symmetric threshold values considered. However, the X^2 statistics vary between 2.000 and 18.000 and are hence well above reasonable

critical values. The empirical kurtosis of the latent variables is about 66. The bias reduces for higher values of N and m except for the .1/.8/.1 case where the above mentioned (absolute) overestimation occurs for distributions having kurtosis between 4 and 8. For $N = 11$ and $m = 20$ the empirical kurtosis is about 3.4 the absolute bias of the polychoric correlation is less than 0.008 and yet the X^2 values vary between 17 and 152.

5 Summary and Conclusions

Multivariate normality plays a central role in estimating the correlation structure of latent variables underlying observed categorical indicators. In the case of two latent variables the polychoric correlation is applied if both variables are only observed categorically whereas polyserial correlation is used if one latent variable is observed directly. In both cases the normality assumption can be tested using Pearson's X^2 test in the former case and tests based on the sample kurtosis of the continuous variable in the latter case (Johnson et al., 1994 p.169). In many empirical applications the normality assumption is rejected. Lee and Lam (1988) developed ML procedures based on other distributions. One problem that occurs is the a priori choice of the correct distribution.

Our goal in this paper is to analyze the bias resulting if estimation is based on normality when the true distribution is a member of the elliptical symmetric distribution class. The distributions all have a linear regression function if it exists. The simulation results show that for those threshold values where the discrete indicators have equal probabilities the bias of the polychoric correlation is smallest. It increases with higher values of kurtosis and with uneven probabilities of the discrete indicators. As an implication, researchers designing questionnaires using ordered categories should try formulating the question so that almost equal probabilities will result. This can be achieved by suggesting thresholds. For instance in business surveys, if a monthly change in a variable is of interest one can suggest that the "unchanged" category is about $\pm a\%$ change in the latent variable with a being appropriately chosen. The more mass is in the middle category, the higher the bias if the true distribution is not normal. High frequencies in the middle category can for instance be observed for most questions in the monthly business survey of the *ifo Institute*, Munich, and in the quarterly business survey of the Centre for European Economic Research, Mannheim. However, even for threshold values implying uneven frequencies of the categories the bias is small for a wide range of elliptical distributions.

For some platykurtic distributions we observe low X^2 test statistics suggesting no deviation from normality. But the bias of the polychoric correlation is already of considerable magnitude. Thus, this test statistic is not useful in finding the most appropriate distribution within the given class since several local minima are present. Again, the polychoric correlation is more robust for platykurtic distributions if even frequencies of the discrete indicator are present.

In the situation where one latent variable is observed directly and the other is observed categorically the polyserial correlation and Brillinger's estimator can be applied. Our results show that Brillinger's estimator is sensitive to deviations from normality whereas the polyserial correlation is robust to a certain extent. Higher kurtosis leads to higher bias whereas the bias is still negligible for negative excess kurtosis.

Simulations for other values of the correlation parameter, not shown in this paper, yield qualitatively the same results. In general, the simulation study indicates that for any given frequency distribution of the categorical indicators there is a set of elliptically symmetric distributions for which the polychoric (polyserial) correlation is a good approximation for the true correlation parameter.

References

Anderson, T. W., (1984), *An Introduction to Multivariate Statistical Analysis*, 2nd edition, John Wiley & Sons, New York.

Brillinger, D.R., (1982), A Generalized Linear Model with Gaussian Regressor Variables, in: Bickel, J.,Doksum, K.A. and J.L. Hodges (Hrsg.), *A Festschrift for Erich L. Lehmann*, Woodsworth International Group, Belmont,CA.

Fang, K.-T., S. Kotz, and K.-W. Ng, (1990), *Symmetric Multivariate and Related Distributions*, Chapman and Hall, London, New York.

Hamdan, M.A., (1970), The Equivalence of Tetrachoric and Maximum Likelihood Estimates in 2×2 Tables, *Biometrika Vol. 57*, 212–215.

Johnson, N., S. Kotz, and N. Balakrishnan, (1994), *Continuous Univariate Distributions* Volume I, 2nd edition, John Wiley & Sons, New York.

Kukuk, M., (1991), *Latente Strukturgleichungsmodelle und rangskalierte Daten*, Hartung–Gorre, Konstanz.

Kukuk, M., (1994), Distributional Aspects in Latent Variable Models, *Statistical Papers, Vol. 35*, 231-242.

Lee, S.-Y. and M.-L. Lam, (1988), Estimation of Polychoric Correlation with Elliptical Latent Variables, *Journal of Statistical Computation and Simulation, Vol. 30*, 173-188.

Mardia, K.V., (1970), *Families of Bivariate Distributions*. Hafner Publ., Darien, Conn.

Olsson, U., (1979), Maximum Likelihood Estimation of the Polychoric Correlation Coefficient, *Psychometrika Vol.44 No.4,* 443–460.

Olsson, U., F. Drasgow and N. Dorans (1982), The Polyserial Correlation Coefficient, *Psychometrika Vol.47 No.3,* 337–347.

Pearson, K., (1901), Mathematical Contributions to the Theory of Evolution. VII. On the Correlation of Characters Not Qualitatively Measurable, *Philosophical Transactions of the Royal Society of London, Series A, Vol.195,* 1–47.

Poon, W.–Y. and S.–Y. Lee, (1987), Maximum Likelihood Estimation of Multivariate Polyserial and Polychoric Correlation Coefficient, *Psychometrika Vol.52 No.3,* 409–430.

Ronning, G. and M. Kukuk, (1996), Efficient Estimation of Ordered Probit Models, *Journal of the American Statistical Association, Vol. 91*, 1120-1129.

Ruud, P.A., (1983), Sufficient Conditions for the Consistency of Maximum Likelihood Estimation Despite Misspecification of Distribution in Multinomial Discrete Choice Models, *Econometrica, Vol. 51*, 225-228.

Ruud, P.A., (1986), Consistent Estimation of Limited Dependent Variable Models Despite Misspecification Of Distribution, *Journal of Econometrics, Vol. 32*, 157-187.

Stoker, T.M., (1986), Consistent Estimation of Scaled Coefficients, *Econometrica, Vol. 54*, 1461-1481.

Tallis, G.M., (1962), The Maximum Likelihood Estimation of Correlation from Contingency Tables, *Biometrics Vol.18,* 342–353.